



An expression tag toolbox for microbial production of membrane bound plant cytochromes P450

Vazquez-Albacete, Dario; Cavaleiro, Ana Mafalda; Christensen, Ulla; Seppälä, Susanna; Møller, Birger Lindberg; Nørholm, Morten H.H.

Published in:
Biotechnology and Bioengineering

DOI:
[10.1002/bit.26203](https://doi.org/10.1002/bit.26203)

Publication date:
2017

Document version
Peer reviewed version

Document license:
[Unspecified](#)

Citation for published version (APA):
Vazquez-Albacete, D., Cavaleiro, A. M., Christensen, U., Seppälä, S., Møller, B. L., & Nørholm, M. H. H. (2017). An expression tag toolbox for microbial production of membrane bound plant cytochromes P450. *Biotechnology and Bioengineering*, 114(4), 751-760. <https://doi.org/10.1002/bit.26203>

An expression tag toolbox for microbial production of membrane bound plant cytochromes P450[†]

Dario Vazquez-Albacete¹, Ana Mafalda Cavaleiro¹, Ulla Christensen¹, Susanna Seppälä¹, Birger Lindberg Møller^{3,4} and Morten H. H. Nørholm^{1,4}

¹ Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle allé 6, Hørsholm, Denmark; ³ Plant Biochemistry Laboratory, Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, Frederiksberg C, Copenhagen, Denmark; ⁴ Center for Synthetic Biology: bioSYNergy, University of Copenhagen, Thorvaldsensvej 40, Frederiksberg C, Copenhagen, Denmark.

Address correspondence to MHHN, morno@biosustain.dtu.dk. Phone: +45 217-99184 Fax: +45-353-33300

[†]This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/bit.26203]

Additional Supporting Information may be found in the online version of this article.

This article is protected by copyright. All rights reserved

Received August 8, 2016; Revision Received October 7, 2016; Accepted October 10, 2016

This article is protected by copyright. All rights reserved

Abstract

Membrane-associated Cytochromes P450 (P450s) are one of the most important enzyme families for biosynthesis of plant-derived medicinal compounds. However, the hydrophobic nature of P450s makes their use in robust cell factories a challenge.

Here we explore a small library of N-terminal expression tag chimeras of the model plant P450 CYP79A1 in different *Escherichia coli* strains. Using a high-throughput screening platform based on C-terminal GFP fusions, we identify several highly expressing and robustly performing chimeric designs. Analysis of long-term cultures by flow cytometry showed homogeneous populations for some of the conditions.

Three chimeric designs were chosen for a more complex combinatorial assembly of a multigene pathway consisting of two P450s and a redox partner. Cells expressing these recombinant enzymes catalysed the conversion of the substrate to highly different ratios of the intermediate and the final product of the pathway. Finally, the effect of a robustly performing expression tag was explored with a library of 49 different P450s from medicinal plants and nearly half of these were improved in expression by more than 2-fold. The developed toolbox serves as platform to tune P450 performance in microbial cells, thereby facilitating recombinant production of complex plant P450-derived biochemicals. This article is protected by copyright.

All rights reserved

Keywords: cytochromes P450, terpenoids, membrane protein, gene expression, N-terminal tag, medicinal plants

Introduction

Cytochromes P450 (P450s) are ubiquitous in plants, where they are involved in the biosynthesis of a wide range of bioactive secondary metabolites. These enzymes have attracted scientific and commercial interest because they modify complex carbon backbones in a highly stereo- and regiospecific manner. Plant P450s are typically membrane-bound by an N-terminal transmembrane helix and depend on a properly incorporated heme group in the catalytic domain. These features make plant P450s a challenge when it comes to recombinant production. Through a myriad of tailoring reactions P450s catalyze the last steps that lead to functionalization of many relevant medicinal compounds (Podust and Sherman 2012). Terpenoids are an example of such naturally occurring compounds, which have shown promising pharmacological activity (King et al. 2014). Unfortunately terpenoids and similar specialized metabolites are often produced in scarce amounts *in planta* making direct extraction from feedstocks economically unsustainable. Taxol and artemisinin are prominent examples of terpenoids used in treatment of cancer and malaria, respectively; both compounds are produced by medicinal plants through complex metabolic pathways involving P450s (Ajikumar et al. 2010; Chang et al. 2007).

Heterologous production of eukaryotic proteins in microbes such as *Escherichia coli* often requires a tedious optimization process. Hydrophobic membrane proteins, such as P450s, are particularly challenging as they tend to form insoluble inclusion bodies (Schlegel et al. 2010). High-throughput screening methods allow for fast design-build-test cycles through identification of highly expressing constructs, and reporters such as the green fluorescent protein from *Aequorea victoria* or bacteriorhodopsin

from *Haloarcula marismortui* can enable quick assessment of properly targeted integral membrane proteins (Drew et al. 2006; Hsu et al. 2013).

A common strategy to overcome challenges with membrane protein production consists of *E. coli* strains equipped with promoters suitable for subtle tuning of expression. For example, the KRX, the Lemo21(DE3) or the Walker strains C41 and C43 are known for higher tolerance for protein production (Giacalone et al. 2006; Miroux and Walker 1996; Wagner et al. 2008). In the KRX strain, a rhamnose-tunable promoter drives expression of a genome-integrated T7 RNA polymerase, whereas in Lemo21(DE3) heterologous gene transcription is tuned by rhamnose-titrated expression of the T7 RNA polymerase inhibitor T7 lysozyme (Schlegel et al. 2013). The Walker strains were evolved to tolerate high-level expression of a toxic membrane protein (Miroux and Walker 1996).

In addition to engineered strains, a handful of strategies have been attempted to specifically facilitate P450 expression in bacterial systems. These include truncation or replacement of the native membrane anchor to solubilize the enzyme, and replacement of the plant membrane-targeting signal with a peptide better recognized by the bacterial membrane translocation machinery (Mergulhao et al. 2005; Pritchard et al. 1997; Sibbesen et al. 1995; Smith et al. 2007). A popular strategy employs the engineered N-terminal sequence from the bovine P450 17 α -hydroxylase, which has been used to produce mammalian and plant P450s in microbes (Bak et al. 1998; Barnes et al. 1991). For example the engineering of several N-terminal P450 anchors resulted in higher artemisinin-precursor production in *E. coli*, showing the positive impact of N-terminal modifications for the complex biosynthesis of a valuable chemical (Chang et al. 2007).

Although a number of N-terminal modifications are available for P450 tailoring, the toolbox for doing so is still limited, especially considering the enormous diversity of plant P450s and complex metabolic pathways. Thus, expanding the number of N-terminal tags, benchmarking the performance of such standard biological parts and testing their compatibility with a large number of e.g. plant P450s sequences could be of great value.

Here, we explore a number of different N-terminal tags for microbial P450 production using the well-characterized *Sorghum bicolor* CYP79A1 as a model plant P450 (Nielsen and Moller 2000). Further, we explore the effect of these peptides in different microbial strains and on a single cell level by flow cytometry. Finally, we demonstrate the use of the N-terminal tag toolbox for subtle expression of a more complex multigene P450 pathway and with a library of 49 different medicinal P450s.

Materials and Methods

Strains and growth conditions

For cloning NEB® 5-alpha Competent *E. coli* cells (New England Biolabs, MA, USA) were transformed and propagated in standard LB media as described by the manufacturer. Chemically competent *E. coli* strains KRX (Promega, Madison, USA) and One Shot® BL21(DE3) (Invitrogen, CA, USA) bearing the pLemo plasmid (referred as strain Lemo21(DE3)), were transformed with the corresponding pET28a(+)-derived constructs (Mirzadeh et al. 2015b; Wagner et al. 2008). Cells were grown in Terrific Broth (TB) and, for Lemo21(DE3), 0.5, 1 and 5 mM of rhamnose (Sigma-Aldrich, St. Louis, USA) was supplemented in the medium. Three biological replicates of each construct were inoculated in TB media supplemented with 50 µg/mL Kanamycin and 25 µg/mL Chloramphenicol and grown over-night.

Expression cultures were initiated at a final OD of 0.05 in 200 µl of fresh TB medium in 96-well cell culture microtiter plates. Absorbance measurements were carried out at 600 nm in a SynergyMx, SMATLD plate reader (BioTek, Winooski, USA). Growth was performed for approximately 2 h at 30°C, 250 rpm in an Innova®44R incubator shaker system (5 cm orbital shaking) (New Brunswick Scientific, Eppendorf, USA) upon reaching 0.5 optical density. Protein expression was induced with IPTG (isopropyl β-D-1-thiogalactopyranoside, dioxane free, Thermo Scientific, Waltham, USA) at a final concentration of 0.4 mM and rhamnose 5 mM for KRX strain (Sigma-Aldrich, MO, USA). Expression cultures were grown at 25°C for 24 h. shaking.

PCR and uracil excision cloning

Uracil excision cloning with PCR-amplified fragments was used to assemble all DNA constructs as described previously (Cavaleiro et al. 2015). All oligonucleotides used are listed in Table S1. PCR products were purified from 1% agarose gels using NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, Düren, Germany). Purified PCR products were mixed with 1 µL USER™ enzyme (New England BioLabs, Ipswich, USA) for 30 min at 37°C and at the *T_m* of the respective primer overhangs for 15 min. DNA was quantified in Nanodrop spectrophotometer 2000 instrument (Thermo Scientific, Waltham, USA). Transformations were carried out in chemically competent NEB5α cells with 5 µL of the uracil excision reaction. Transformants were screened by colony PCR with OneTaq 2X Master mix (New England BioLabs, Ipswich, USA). for plasmid isolation with a QIAprep Spin Miniprep Kit (Qiagen) and sequenced according to the service supplier instructions (Eurofins Genomics, Ebersberg, Germany). All resulting plasmid constructs are listed in Table S2.

Generation of N-terminally tagged P450 libraries

A library of different synthetic N-terminal tags for the cytochrome *SbCYP79A1* was constructed by uracil excision cloning as described above. A pET28a(+) expression vector bearing the native sequence of the *Sorghum bicolor* cytochrome *SbCYP79A1*, a TEV protease cleavage site, GFP folding reporter and His-tag tail (*pET-CYP79A1*) was used as template to insert the different N-terminal tags (Drew et al. 2001). A version of the same plasmid with the *SbCYP79A1* lacking the amino acids one to 35 (*tSbCYP79A1*) was also prepared. The genes *ppiD*, *yfgM*, *yfdG*, *yhaH*, *ymcD*, *gltS*, *tyrP* and the 28aa tag sequence were amplified by PCR with the oligonucleotides (L4-L17) from *E. coli* MG1665 genome and cloned by two-fragment uracil excision cloning into the *pET-CYP79A1* backbone linearized with oligonucleotides L2 and L3. The predicted transmembrane region (residues 1-39) of *zipA* and the two predicted transmembrane helices of the *ycjF* (residues 1-147) were cloned the same way into the *pET-tCYP79A1* plasmid linearized with oligonucleotides L1 and L3. Finally, signal peptides from *dsbA*, *ompA*, *malE* and *nlpA* with the lipid anchoring sequence CDQSSS (*nlpA_as*) and without this sequence (*nlpA_ns*) were cloned into the *pET-CYP79A1* plasmid by one-fragment uracil excision cloning using the oligonucleotides L22 to L34. Predictions of transmembrane regions from endogenous *E.coli* membrane proteins were performed with TMHMM v. 2.0 and TOPCONS (Hessa et al. 2007; Tsirigos et al. 2015). Signal peptides were predicted using SignalP 4.0 server (Petersen et al. 2011).

Additionally, a library of 49 codon optimized P450-coding genes from *Euphorbia peplus* (*Ep*), *Coleus forskohlii* (*Cf*), *Heliconius Melpomene* (*Hm*), *Thapsia garganica* (*Tg*), *Thapisa villosa* (*Tv*) and *Tripterygium willfordii* (*Tw*) was purchased from GeneArt™ (Thermo Fisher Scientific, Wilmington, USA). ORFs were inserted in the

pET28a(+) vector flanked by restriction sites *XhoI* and *BamHI* (Table S3). Oligonucleotides L49 and L50 used to PCR amplify the pET28a(+) backbone had 5'-end *XhoI* and *BamHI* restriction sites. Plasmids were digested with Fast Digest™ *XhoI* and *BamHI* (Thermo Fisher Scientific, Wilmington, USA) for 20 min at 37°C, 5 min at 80°C and 10 min at 4°C. Ligation was performed with T4 DNA ligase (Thermo Fisher Scientific, Wilmington, USA) following specifications of the manufacturer. Mixtures were kept on ice prior to transformation.

Dhurrin pathway assembly

The N-terminal tags *dsbA*, *ompA* and the 28aa tag sequence versions of the cytochrome *SbCYP71E1* were cloned as described for the *SbCYP79A1* in section 2.2 with oligonucleotides L35 to L40. For assembly of the multigene pathway, two-fragment uracil excision was performed; tagged and non-tagged versions of the two cytochromes without the GFP moiety were cloned in all possible combinations as two separate transcriptional units with the oligonucleotides L41 to L44. See list of resulting constructs (Table S2).

A cytochrome reductase from *Sorghum bicolor* (*CPR*) containing a Strep-HRV3C tag in the 5' end was also cloned by two-fragment uracil excision into the pET28a(+) vector from which the GFP folding reporter was removed. The origin of replication and the antibiotic resistance marker were swapped by the corresponding pSEVA63 parts to avoid plasmid incompatibility (Durante-Rodriguez et al. 2014).

GFP measurements and expression levels

Whole cell fluorescence was measured in a SynergyMx SMATLD plate reader (BioTek, Winooski, USA) from 96-well cell culture plates within the 485-512 nm range. To estimate expression levels, purified GFP was used as standard, diluted in cultured cells transformed with the empty vector (negative control) in order to simulate the quenching effect as described previously (Mirzadeh et al. 2015a). Single cell measurements were also performed in a BD LSRFortessa™ flow cytometer equipped with HTS module (Becton, Dickinson and Company, USA). Cells were diluted 1/200 in PBS, incubated for 15 min at room temperature and green fluorescence was collected in the FITC channel (488 nm emission wavelength). Expression thresholds were determined with controls consisting of cells containing the empty vector and soluble GFP expressing cells for both tested *E. coli* strains. Flow cytometry data was exported in FCS files and analysed using the *flowViz* and *flowcore* packages with *R* software (Hahne et al. 2009; Sarkar et al. 2008) (<http://www.R-project.org/>)

Preparation of *E. coli* membrane fractions

In order to recover the overexpressed membrane-associated cytochromes, *E. coli* membranes were isolated as previously described (Drew et al. 2006). Briefly, expression was carried out using 50 ml in flasks and harvested after 24 h as for the previous experiments. Cells were lysed by mechanical disruption using an Emulsiflex C-50 (Avestin Europe GmbH) instrument at 15,000-20,000 p.s.i. at 4°C. Unbroken cells were removed by centrifugation at 8,000 xg, 10 min at 4°C and the supernatant collected for further fractionation. Membranes were separated by ultracentrifugation at 150.000 xg, 45 min, 4°C and recovered from the pellet in 20 mM Tris buffer

pH=7.8, 20% glycerol. The supernatant containing the cytoplasmic fraction was collected and concentrated 5X using Amicon Ultra centrifugal filter units MWCO 10 KDa (Millipore, Billerica, MA, USA). Cell fractions were snap-frozen in liquid nitrogen and stored at -80°C upon analysis. Protein concentration was determined using BCA protein assay (Thermo Scientific, USA) prior to storage at -80°C.

SDS-PAGE analysis

Membrane-associated proteins were identified by the presence of full length fluorescent bands in the membrane preparations. Membrane proteins were adjusted to the same concentration and analyzed by standard SDS-PAGE loading 100 µg of total protein of each sample in 10% acrylamide gels. Fluorescence was detected with a G:BOX Bioimager UV-table (Syngene, Cambridge, UK).

Activity assay of dhurrin pathway cytochrome P450 enzymes and detection by Thin Layer Chromatography (TLC)

The activity of the P450s in the dhurrin pathway was analyzed qualitatively as previously reported with minor modifications (Kahn et al. 1999). *E. coli* cells transformed with the two-cytochrome pathway and the CPR were cultured accordingly in 1 ml volume. After 24 h expression, absorbance was measured; cells were washed twice with 50 mM KPi buffer (Potassium phosphate) and adjusted to the same optical density units (ODU). The pathway activity was assayed *in vivo* in a reaction volume of 15 µL consisting of 10 µL of cell suspension (ODU=1), 5 mM NADPH, 0.1 mM L-tyrosine (Sigma-Aldrich, St. Louis, USA), 1.5 µL of [U-¹⁴C]tyrosine (0.05 mCi, 482 mCi/mmol) and adjusted with 50 mM KPi buffer. The reaction was performed for 60 min at 30°C and 700 rpm in a benchtop shaker (Grant Instruments, Cambridge, UK). The products of the pathway, (*E*)-*p*-

hydroxyphenylacetaldoxime and p-hydroxymandelonitrile (oximes), were extracted from the reaction by addition of one volume of ethyl acetate (Sigma-aldrich, MO, USA). The mixture was centrifuged to remove cell debris at 16,000 xg for 10 min and the organic phase transferred to a silica gel plate (Sigma-Aldrich, St. Louis, USA). The extract was eluted in a glass chamber with a mobile phase composed by toluene, ethyl acetate and methanol (30:8:1). Radioactive products were visualized in a Cyclone Plus Phosphor Imager instrument (Perkin Elmer, Massachusetts, USA). Relative intensity of radioactive oximes was determined by *ImageJ* (U. S. National Institutes of Health, Bethesda, Maryland, USA, <http://imagej.nih.gov/ij/>).

Results

Expression analysis of a library of N-terminal tags of *SbCYP79A1*

We constructed a synthetic N-terminal tag library for the plant cytochrome P450 CYP79A1. The N-terminal tags were cloned as CYP79A1 fusions in a pET28a(+) expression vector. The vector contains a T7 promoter upstream of the gene, a TEV protease cleavage site at the 3' end of the gene followed by a GFP folding reporter and an octahistidine tag (Norholm et al. 2013). The N-terminal tags were chosen based on their topologies and expression profiles determined in other studies (Daley et al. 2005; Lee et al. 2013; Sletta et al. 2007), membrane localization in *E. coli* (Gotzke et al. 2015) or relaxation of mRNA folding energy (Kudla et al. 2009). The N-terminal tag-types fall in five categories (Fig. 1A): 1) bacterial membrane anchors with the C-terminal facing the cytoplasmic side (C-in; SohB, LepA, ZipA and YcjF), 2) proteins with the C-terminal facing the periplasmic side (C-out; YfdG, YhaH, YmcD, PpiD and YfgM), 3) signal peptides (MalE, OmpA, DsbA and NlpA_{as} carrying a lipid anchoring sequence (as) and NlpA_{ns} without lipid anchoring sequence), 4)

transporters (GltS and TyrP) and 5) a previously described expression-enhancing peptide (28aa tag) (Norholm et al. 2013). The C-in tags and transporters replaced the transmembrane segment of the P450; whereas in the other constructs the P450 transmembrane segment was left intact.

The N-terminal tag chimeric constructs of CYP79A1 were transformed into the KRX *E. coli* strain that harbor a T7 RNA polymerase under the control of the tunable rhamnose promoter. Clones were grown on TB medium and induced with both rhamnose and IPTG for 24 hours. Pilot experiments showed no major effects of varying the rhamnose concentration and thus we used a fixed rhamnose concentration of 5 mM similar to previous work (Locatelli-Hoops et al. 2013). Single cell and whole cell fluorescence were subsequently measured as indication of the presence of CYP79A1 chimeras. The data shows no fluorescence of the wild type *cyp79a1* gene fusion in *E. coli*, whereas all N-terminal tagged chimeras were produced to a detectable level (Fig. 1B). Most of the chimeras under the C-in category, which required removal of the CYP79A1 transmembrane segment, exhibited the strongest fluorescent signal together with the 28aa tag chimera. Yields were estimated based on levels of GFP and were up to 40 mg/L for the best performing constructs. Chimeras based on the C-out category and signal peptides were less fluorescent than the C-in constructs, with the YfgM chimera performing best at 15 mg GFP/L. The presence of full-length chimeras in membrane fractions was confirmed for highest expressed constructs by cell fractionation and SDS-PAGE and free GFP was observed in the cytoplasmic fraction, indicating partial degradation of the fusion proteins (Fig. S1).

Next, the CYP79A1 chimeras were tested in the *E. coli* B strain Lemo21(DE3) (Wagner et al. 2008). Furthermore, to find the best expression conditions for each chimera considering maximum GFP signal and homogeneity of cell populations, the

cells were induced with three different rhamnose concentrations. As in KRX, the wild type *cyp79a1* gene was poorly expressed and expression was significantly improved by nearly all N-terminal tags (Fig. 2A). While in the KRX strain the highest fluorescence was observed for C-in chimeras, in the Lemo21(DE3) strain, the highest signal was obtained for the PpiD tag (C-out) in the absence of rhamnose, followed by the signal peptide chimeras and the 28aa tag at 0.5 mM rhamnose, with values ranging from 80 to 100 mg GFP/L (Fig. 2A). Different constructs responded differently to rhamnose induction. As judged from GFP-levels, cells expressing signal peptide- and 28aa tag chimeras performed best with the lowest concentration inducer (0.5 mM rhamnose), whereas cells expressing most of the other chimeras responded negatively to any presence or increase in rhamnose concentration, corroborating the notion that inducer levels should be titrated for each construct (Fig. 2A). For the signal peptide- and 28aa tag-chimeras, in the absence of rhamnose, GFP fluorescence was lost after 24 hours, whereas the addition of 0.5 mM rhamnose appears not only to give higher expression, but, importantly, fluorescence is maintained after 24 hours in homogenous populations. This demonstrates the value of expression titration for increasing phenotype stability in this strain.

Taken together, these results showcase the highly different performance of different N-terminal tags on balancing expression of a model P450. Signal peptides and the 28aa tag showed high fluorescence of P450 chimeras in both strains. Flow cytometry analysis showed loss of long term expression with some chimeric constructs and subtle expression tuning helped maximizing and stabilizing the phenotype.

Multigene pathway activity

Dhurrin is a natural compound synthesized from the amino acid tyrosine and released by plants to deter herbivores from feeding on leaves. The first two steps in the dhurrin pathway of *Sorghum bicolor* are catalyzed by the cytochromes P450 CYP79A1 and CYP71E1, respectively. The two intermediates generated by each enzyme, (*E*)-*p*-hydroxyphenylacetaldoxime and (*E*)-*p*-hydroxymandelonitrile, can be easily detected and separated from the substrate tyrosine by thin layer chromatography (TLC) (Kahn et al. 1999). The catalytic reaction is assisted by the electron donor cytochrome P450 reductase (*SbCPR2b*) (Laursen et al. 2011). Seeing the robust performance of the OmpA, DsbA and 28aa tagged CYP79A1 chimeras, we expanded our studies by systematically combining chimeric versions of the dhurrin pathway P450s in the same expression vector as two separate transcriptional units consisting of a T7P promoter combined with a T1 terminator for the CYP79A1 chimeras or a T7 terminator for CYP71E1 chimeras (Fig. 3A and Fig. S2A). The vectors harboring the P450s and the partnering reductase (CPR) ORFs were co-transformed and expressed for 24 h in the KRX strain since it displayed superior expression stability and has minor basal expression of the T7 RNA polymerase (Ishido et al. 2011). The activity of the pathway was subsequently assayed by addition of radioactively labelled tyrosine to cultures normalized to the same optical density units (ODU). The two oximes produced by CYP79A1 and CYP71E1 were extracted from the reaction with ethyl acetate and analyzed by TLC (Fig. 3B and Fig. S2B). Cells producing a combination of 28aa tagged *SbCYP79A1* and wild type *SbCYP71E1* produced only the first intermediate, (*E*)-*p*-hydroxyphenylacetaldoxime at a high level (Fig. 3B). When both P450s are N-terminally tagged, the (*E*)-*p*-hydroxymandelonitrile is produced and the combination of a 28aa tagged *SbCYP79A1* and an OmpA-tagged *SbCYP71E1*

yielded the highest level of the final product of the pathway. As expected from the low expression level, no combinations with the wild type cytochrome *cyp79a1* generated any product. These results demonstrate the usefulness of an extended N-terminal tag toolbox for cell factory design.

Expression of a library of tagged medicinal P450s

To test the broader applicability of the N-terminal tagging strategy, we next synthesized a library of 49 codon harmonized P450 genes derived from the medicinal plants *Euphorbia peplus* (*Ep*), *Coleus forskohlii* (*Cf*), *Thapsia garganica* (*Tg*), *Thapsia villosa* (*Tv*), *Tripterygium willfordii* (*Tw*) and the insect *Heliconius Melpomene* (*Hm*) that has metabolic pathways and P450s that are related to plant enzymes (Chauhan et al. 2013). The genes were produced with and without an N-terminal 28aa tag, all fused to a C-terminal GFP as described above. The fold change in fluorescence between the wild type and chimeric P450s was used to estimate expression level differences. The 28aa tag improved expression of 68% of the P450 genes and expression of 44% of the library was increased two-fold or more with the 28aa tag. These plants are highly relevant since they are involved in the biosynthesis of industrially relevant compounds such as diterpenes. Two of the most impressive improvements were observed for the *EpCYP76A1* with a 35 fold and the *EpCYP726A1* with a 9 fold increase. The latter is involved in the production of ingenol, a relevant compound for the treatment of actinic keratinosis (Jorgensen et al. 2013; King et al. 2014).

Discussion

Membrane-associated cytochromes P450 are important biocatalysts in the biosynthesis of natural compounds such as terpenoids (Podust and Sherman 2012). A major bottleneck for the production of functionalized terpenoids in microbial hosts is robust expression (Chang and Keasling 2006). Here, we assayed the performance and stability of *E. coli* cultures expressing N-terminally modified plant P450 enzymes. Using a GFP-based platform that allows simple assessment of expression levels, membrane localization and protein integrity, we show that modifying the plant enzymes with signal peptides or protein tags derived from the microbial host significantly increase levels of the chimeric proteins, suggesting that host-tailored modifications of the target protein may be key to heterologous protein production. We also specifically verified the beneficial effect on gene expression of a 28 amino acid tag that has been shown to boost expression, possibly by relaxing mRNA structure (Kudla et al. 2009). Further, we compared the performance of the two different *E. coli* strains KRX and Lemo21(DE3) (Giacalone et al. 2006; Miroux and Walker 1996; Wagner et al. 2008) - both strains allow for tunable gene expression and have previously shown promise for the production of challenging proteins. Whole cell fluorescence varied between the strains, with Lemo21(DE3) strain outperforming KRX in terms of total estimated protein yield. However, this pattern may change upon tuning of T7 polymerase expression with rhamnose in Lemo21(DE3). Analysis of cell fractions demonstrates the presence of full-length enzymes in the membrane fractions. Unfortunately the level of full-length enzyme does not directly correlate with whole cell fluorescence, likely due to the presence of free soluble GFP in the cytoplasmic fraction. Despite this drawback the platform provides a valuable screening scheme to

discriminate between highly or poorly expressed chimeric variants, facilitating further optimization efforts.

Different N-terminal tag chimerae showed subtle expression differences in the two strains and with different rhamnose levels in Lemo21(DE3). For example, using flow cytometry, with the signal peptide base constructs in the Lemo21(DE3) strain, 0.5 mM rhamnose stabilized the homogeneity and maintained expression at high levels. In contrast, absence of rhamnose led to loss of fluorescence in several cultures over time. This is likely due to over-expression of the P450 saturating the Sec translocon machinery, which may result in protein aggregation (Klepsch et al. 2011). The KRX strain displayed superior expression stability, which could be related to a lower basal expression of the T7 RNA polymerase (Ishido et al. 2011). For this reason KRX was selected as preferred strain for further experiments. The two rhamnose-based expression strains had no or minimal effects on increasing fluorescence of the wild type gene *cyp79a1* fusion. Therefore it is clear that an N-terminal tag library provides important tools to unblock expression bottlenecks in *E. coli*.

We further demonstrated the impact of combinatorial chimeric constructs on pathway optimization and microbial production of the plant-derived chemical dhurrin. Two P450s participate in the formation of dhurrin from tyrosine and the P450s receive electrons from an associated oxidoreductase. Using the best overall performing expression tags we assembled different combinations of the dhurrin pathway as separate transcriptional units in *E. coli*, and assayed for the formation of the intermediate (*E*)-*p*-hydroxyphenylacetaldoxime and product (*E*)-*p*-hydroxymandelonitrile. Only when the P450s were tagged, product formation was detected, but the different combinations produced highly different amounts of the final product of the pathway. This emphasizes the usefulness of an extended toolbox

of different N-terminal tags, to serve as platform for biosynthetic pathway engineering.

We further analyzed the effect of the 28aa tag on a library of 49 P450 genes from five medicinal plant species and one insect. These coding sequences were harmonized for codon usage in *E. coli*, but many factors determine the success of codon optimization strategies (Menzella 2011). Our results show that a large number of P450s from different species are positively affected by N-terminal tags even when they are all codon optimized for the same organism. The most dramatic improvement was observed for *EpCYP76A1*, which has not previously been expressed in a microbial host and other members of this family were also well-expressed. This is exciting, because *Arabidopsis thaliana* and *Catharantus roseus* homologs of the CYP76 family have been associated with geraniol and hydroxygeraniol hydroxylase activities - key enzymes in the early steps in the biosynthesis of valuable chemicals such as the anticancer molecule vinblastine (Hofer et al. 2013). Further, we observed a 9-fold expression increase for *EpCYP726A1*, which is involved in the biosynthesis of important diterpenoids such as ingenol, a commercially available medicinal compound currently produced by a complex 9-step chemical reaction (Jorgensen et al. 2013; King et al. 2014). Future production of these medicinal compounds in bacteria may be assisted by the aid of the N-terminal tag toolbox described here.

In this study we have found and characterized N-terminal tags that allow or fine-tune expression of functionally active plant P450s in multigene biosynthetic pathways. We further demonstrate the use of these tags in two commonly used *E. coli* protein production strains, test their performance on the single-cell level and a with large medicinal P450 library. These results provide a framework for the broad application of N-terminal tags in the design of microbial cell factories.

This article is protected by copyright. All rights reserved

Declarations

List of abbreviations

P450: Cytochrome P450s

GFP: green fluorescent protein

CYP79A1: Cytochrome P450 CYP79A1

CYP79E1: Cytochrome P450 CYP79E1

CPR: Cytochrome P450 Reductase

28aa tag: short 28 amino acid tag

TLC: Thin Layer Chromatography

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

This work was supported by The Novo Nordisk Foundation and a PhD grant from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme [FP7-People-2012-ITN], under grant agreement No. 317058, "BACTORY". SS is the recipient of VILLUM Foundation's Young Investigator Programme grant VKR023128.

Authors' contributions

Dario Vazquez designed the experiments, carried out cloning of N-terminal tag library, optimized the expression platform and analyzed membrane proteins. Mafalda Cavaleiro designed cloning strategy of P450 libraries, cloned a significant number of constructs and participated in the expression experiments. Ulla Christensen designed the initial expression platform and provided guidance during the study. Susanna Seppälä designed N-terminal tag chimeras and provided guidance during the study.

Birger Lindberg Møller provided insights into biochemistry and diversity of P450 enzymes. Morten H.H. Nørholm proposed, supervised and provided guidance of the study.

Acknowledgements

Björn Hamberger, Søren Bak and Henrik Toft Simonsen are thanked for assistance with selecting P450s for the expression library. Acknowledgement is given to João Cardoso for helping with the R code.

References

- Ajikumar PK, Xiao WH, Tyo KE, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G. 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* 330(6000):70-4.
- Bak S, Kahn RA, Nielsen HL, Moller BL, Halkier BA. 1998. Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from *Sorghum bicolor* (L.) Moench by a PCR approach and identification by expression in *Escherichia coli* of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. *Plant Mol Biol* 36(3):393-405.
- Barnes HJ, Arlotto MP, Waterman MR. 1991. Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in *Escherichia coli*. *Proc Natl Acad Sci U S A* 88(13):5597-601.
- Cavaleiro AM, Kim SH, Seppala S, Nielsen MT, Norholm MH. 2015. Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. *ACS Synth Biol* 4(9):1042-6.
- Chang MC, Eachus RA, Trieu W, Ro DK, Keasling JD. 2007. Engineering *Escherichia coli* for production of functionalized terpenoids using plant P450s. *Nat Chem Biol* 3(5):274-7.
- Chang MC, Keasling JD. 2006. Production of isoprenoid pharmaceuticals by engineered microbes. *Nat Chem Biol* 2(12):674-81.
- Chauhan R, Jones R, Wilkinson P, Pauchet Y, Ffrench-Constant RH. 2013. Cytochrome P450-encoding genes from the *Heliconius* genome as candidates for cyanogenesis. *Insect Mol Biol* 22(5):532-40.
- Daley DO, Rapp M, Granseth E, Melen K, Drew D, von Heijne G. 2005. Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* 308(5726):1321-3.
- Drew D, Lerch M, Kunji E, Slotboom DJ, de Gier JW. 2006. Optimization of membrane protein overexpression and purification using GFP fusions. *Nat Methods* 3(4):303-13.
- Drew DE, von Heijne G, Nordlund P, de Gier JW. 2001. Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. *FEBS Lett* 507(2):220-4.
- Durante-Rodriguez G, de Lorenzo V, Martinez-Garcia E. 2014. The Standard European Vector Architecture (SEVA) plasmid toolkit. *Methods Mol Biol* 1149:469-78.
- Giacalone M, Gentile A, Lovitt B, Berkley N, Gunderson C, Surber M. 2006. Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system. *BioTechniques* 40(3):355-364.
- Gotzke H, Muheim C, Altelaar AF, Heck AJ, Maddalo G, Daley DO. 2015. Identification of putative substrates for the periplasmic chaperone YfgM in *Escherichia coli* using quantitative proteomics. *Mol Cell Proteomics* 14(1):216-26.
- Hahne F, LeMeur N, Brinkman RR, Ellis B, Haaland P, Sarkar D, Spidlen J, Strain E, Gentleman R. 2009. flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* 10:106.

- Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH, von Heijne G. 2007. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* 450(7172):1026-30.
- Hofer R, Dong L, Andre F, Ginglinger JF, Lugan R, Gavira C, Grec S, Lang G, Memelink J, Van der Krol S and others. 2013. Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early steps of the (seco)iridoid pathway. *Metab Eng* 20:221-32.
- Hsu MF, Yu TF, Chou CC, Fu HY, Yang CS, Wang AH. 2013. Using Haloarcula marismortui bacteriorhodopsin as a fusion tag for enhancing and visible expression of integral membrane proteins in Escherichia coli. *PLoS One* 8(2):e56363.
- Ishido T, Yamazaki N, Ishikawa M, Hirano K. 2011. Characterization of DNA polymerase beta from Danio rerio by overexpression in E. coli using the in vivo/in vitro compatible pIVEX plasmid. *Microb Cell Fact* 10:84.
- Jorgensen L, McKerrall SJ, Kuttruff CA, Ungeheuer F, Felding J, Baran PS. 2013. 14-step synthesis of (+)-ingenol from (+)-3-carene. *Science* 341(6148):878-82.
- Kahn RA, Fahrendorf T, Halkier BA, Moller BL. 1999. Substrate specificity of the cytochrome P450 enzymes CYP79A1 and CYP71E1 involved in the biosynthesis of the cyanogenic glucoside dhurrin in Sorghum bicolor (L.) Moench. *Arch Biochem Biophys* 363(1):9-18.
- King AJ, Brown GD, Gilday AD, Larson TR, Graham IA. 2014. Production of bioactive diterpenoids in the euphorbiaceae depends on evolutionarily conserved gene clusters. *Plant Cell* 26(8):3286-98.
- Klepsch MM, Persson JO, de Gier JW. 2011. Consequences of the overexpression of a eukaryotic membrane protein, the human KDEL receptor, in Escherichia coli. *J Mol Biol* 407(4):532-42.
- Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene expression in Escherichia coli. *Science* 324(5924):255-8.
- Laursen T, Jensen K, Moller BL. 2011. Conformational changes of the NADPH-dependent cytochrome P450 reductase in the course of electron transfer to cytochromes P450. *Biochim Biophys Acta* 1814(1):132-8.
- Lee J, Velmurugan N, Jeong K. 2013. Novel strategy for production of aggregation-prone proteins and lytic enzymes in Escherichia coli based on an anchored periplasmic expression system. *Journal of bioscience and bioengineering*.
- Locatelli-Hoops S, Sheen FC, Zoubak L, Gawrisch K, Yeliseev AA. 2013. Application of HaloTag technology to expression and purification of cannabinoid receptor CB2. *Protein Expr Purif* 89(1):62-72.
- Menzella HG. 2011. Comparison of two codon optimization strategies to enhance recombinant protein production in Escherichia coli. *Microb Cell Fact* 10:15.
- Mergulhao FJ, Summers DK, Monteiro GA. 2005. Recombinant protein secretion in Escherichia coli. *Biotechnol Adv* 23(3):177-202.
- Miroux B, Walker JE. 1996. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* 260(3):289-98.

- Mirzadeh K, Martinez V, Toddo S, Guntur S, Herrgard MJ, Elofsson A, Norholm MH, Daley DO. 2015a. Enhanced Protein Production in *Escherichia coli* by Optimization of Cloning Scars at the Vector-Coding Sequence Junction. *ACS Synth Biol*.
- Mirzadeh K, Martinez V, Toddo S, Guntur S, Herrgard MJ, Elofsson A, Norholm MH, Daley DO. 2015b. Enhanced Protein Production in *Escherichia coli* by Optimization of Cloning Scars at the Vector-Coding Sequence Junction. *ACS Synth Biol* 4(9):959-65.
- Nielsen JS, Moller BL. 2000. Cloning and expression of cytochrome P450 enzymes catalyzing the conversion of tyrosine to p-hydroxyphenylacetaldoxime in the biosynthesis of cyanogenic glucosides in *Triglochin maritima*. *Plant Physiol* 122(4):1311-21.
- Norholm MH, Toddo S, Virkki MT, Light S, von Heijne G, Daley DO. 2013. Improved production of membrane proteins in *Escherichia coli* by selective codon substitutions. *FEBS Lett* 587(15):2352-8.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785-6.
- Podust L, Sherman D. 2012. Diversity of P450 enzymes in the biosynthesis of natural products. *Natural product reports* 29(10):1251-1266.
- Pritchard M, Ossetian R, Li D, Henderson C, Burchell B, Wolf C, Friedberg T. 1997. A general strategy for the expression of recombinant human cytochrome P450s in *Escherichia coli* using bacterial signal peptides: expression of CYP3A4, CYP2A6, and CYP2E1. *Archives of biochemistry and biophysics* 345(2):342-354.
- Sarkar D, Le Meur N, Gentleman R. 2008. Using flowViz to visualize flow cytometry data. *Bioinformatics* 24(6):878-9.
- Schlegel S, Klepsch M, Gialama D, Wickstrom D, Slotboom DJ, de Gier JW. 2010. Revolutionizing membrane protein overexpression in bacteria. *Microb Biotechnol* 3(4):403-11.
- Schlegel S, Rujas E, Ytterberg AJ, Zubarev RA, Luirink J, de Gier JW. 2013. Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. *Microb Cell Fact* 12:24.
- Sibbesen O, Koch B, Halkier BA, Moller BL. 1995. Cytochrome P-450TYR is a multifunctional heme-thiolate enzyme catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. *J Biol Chem* 270(8):3506-11.
- Sletta H, Tondervik A, Hakvag S, Aune TE, Nedal A, Aune R, Evensen G, Valla S, Ellingsen TE, Brautaset T. 2007. The presence of N-terminal secretion signal sequences leads to strong stimulation of the total expression levels of three tested medically important proteins during high-cell-density cultivations of *Escherichia coli*. *Appl Environ Microbiol* 73(3):906-12.
- Smith BD, Sanders JL, Porubsky PR, Lushington GH, Stout CD, Scott EE. 2007. Structure of the human lung cytochrome P450 2A13. *J Biol Chem* 282(23):17306-13.
- Tsirigos KD, Peters C, Shu N, Kall L, Elofsson A. 2015. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* 43(W1):W401-7.

Accepted Preprint

Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Høgbom M, van Wijk KJ, Slotboom DJ, Persson JO and others. 2008. Tuning *Escherichia coli* for membrane protein overexpression. *Proc Natl Acad Sci U S A* 105(38):14371-6.

Figures Caption:

Figure 1. Expression screening of N-terminally tagged CYP79A1 chimeras in the *Escherichia coli* KRX strain. **(A)** Illustration of the protein architectures included in the N-terminal tag library with colour coding. Five categories of N-terminal tags are shown in different colours; Green for C-in architectures, blue for C-out architectures, purple for signal peptides, beige for transporters and red for the short 28aa tag sequence. **(B)** Expression levels of N-terminal-tagged chimeras and wild type CYP79A1 (no tag). Expression was induced with IPTG and rhamnose in exponential phase for 24 h and expression levels estimated by whole cell fluorescence in a plate reader. Data represents the average of three biological replicates with standard deviations.

Figure 2. Expression screening of N-terminally tagged CYP79A1 chimeras in the *E. coli* strains Lemo21(DE3) and KRX analysed by whole-cell fluorescence and flow cytometry. **(A)** Expression levels of N-terminal-tagged chimeras and wild type CYP79A1 (non tagged) in the Lemo21(DE3) strain. Gene expression was induced with IPTG and rhamnose (0, 0.5, 1 or 5 mM) in exponential phase for 24 h and whole cell fluorescence was measured in a plate reader. Data represents the average of three biological replicates with standard deviations. Five categories of N-terminal tags are shown in different colours; Green for C-in architectures, blue for C-out architectures, purple for signal peptides, beige for transporters and red for the short 28aa tag sequence. **(B)** Single cell fluorescence measurements by flow cytometry of selected CYP79A1 chimeras expressed in KRX and Lemo21(DE3). Expression was induced for four or 24 h as described above and different constructs were transferred to PBS and fluorescence analysed in the FITC channel. Negative and positive expression

thresholds were determined with control cells containing the empty vector (P1) and cells expressing soluble GFP (P2), respectively.

Figure 3. Activity of the dhurrin pathway assembled and assayed in the *E. coli* KRX strain. **(A)** Left side: Schematic representation of the dhurrin pathway assembled as two separate transcriptional units in the pET28a(+) vector. Combinations of the wild type and N-terminal-tagged cytochromes were assembled into the pET28 vector containing two T7 promoters (T7P) and two different terminators following each ORF. The constructs were co-transformed in *E. coli* with a cytochrome P450 reductase (CPR) construct illustrated on the right side. **(B)** Analysis of activity of the dhurrin pathway P450s by Thin-Layer-Chromatography (TLC). Radioactive spots observed in the lower part of the TLC correspond to (*E*)-*p*-hydroxyphenylacetaldoxime, the enzymatic product of CYP79A1, whereas the spots observed in the upper part correspond to (*E*)-*p*-hydroxymandelonitrile, produced by CYP71E1. Chemical structures are illustrated on the right side of the TLC. Cells were grown in TB media and expression of the multigene pathway was induced for 24 h. Cells were adjusted to the same ODU, fed with [U-¹⁴C]tyrosine and incubated for 1 h. Enzymatic products were extracted from the reaction with ethyl acetate and loaded onto a TLC silica plate. Radioactive products were visualized using phosphor imaging. **(C)** Relative levels of oximes from two biological replicates were quantified in each TLC plate using ImageJ and normalized to the most intense spot.

Figure 4. Average fold change in expression levels between wild type and 28aa tagged P450-encoding genes estimated by GFP fluorescence in *E. coli* KRX. Expression was induced with IPTG and rhamnose in exponential phase for 24 h and

expression levels estimated by whole cell fluorescence in a plate reader. Data represents the average of three biological replicates with standard deviations.

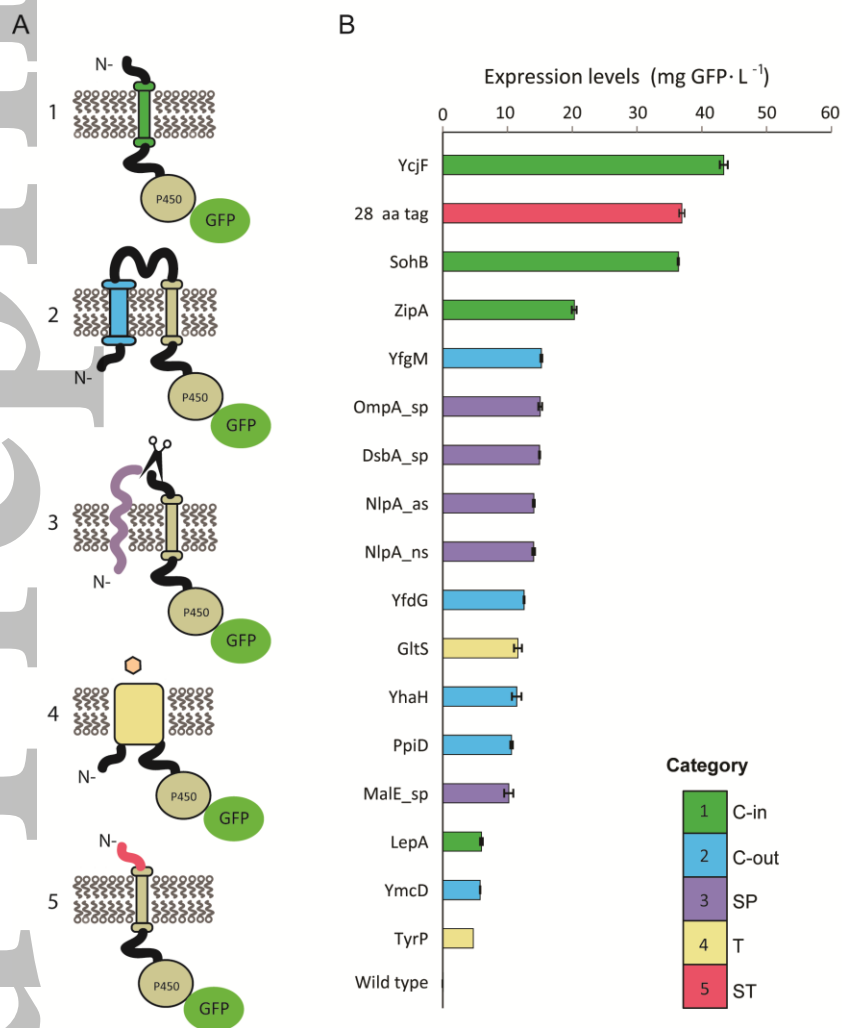


Figure 1

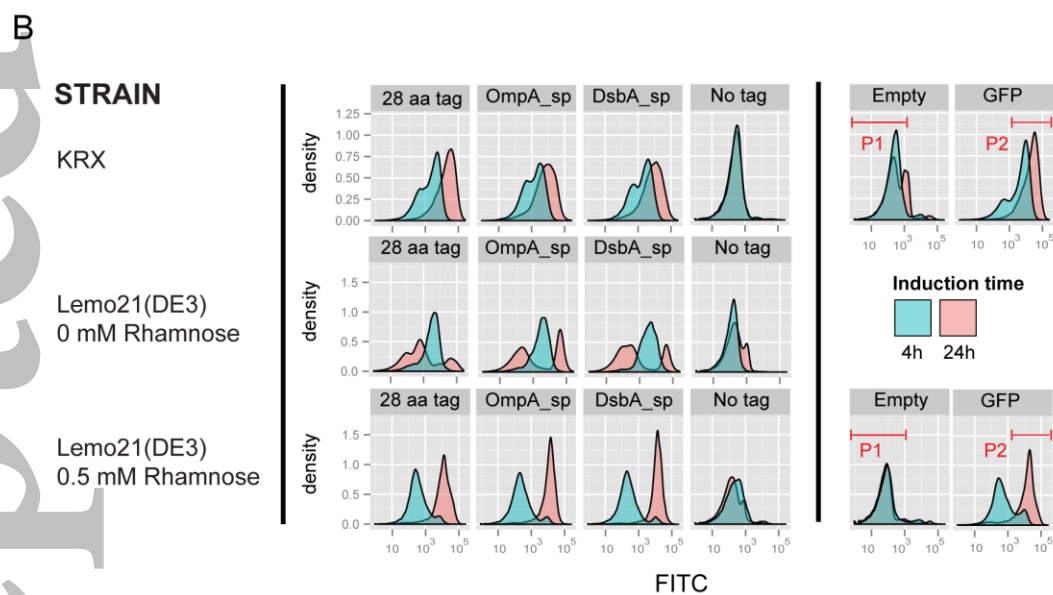
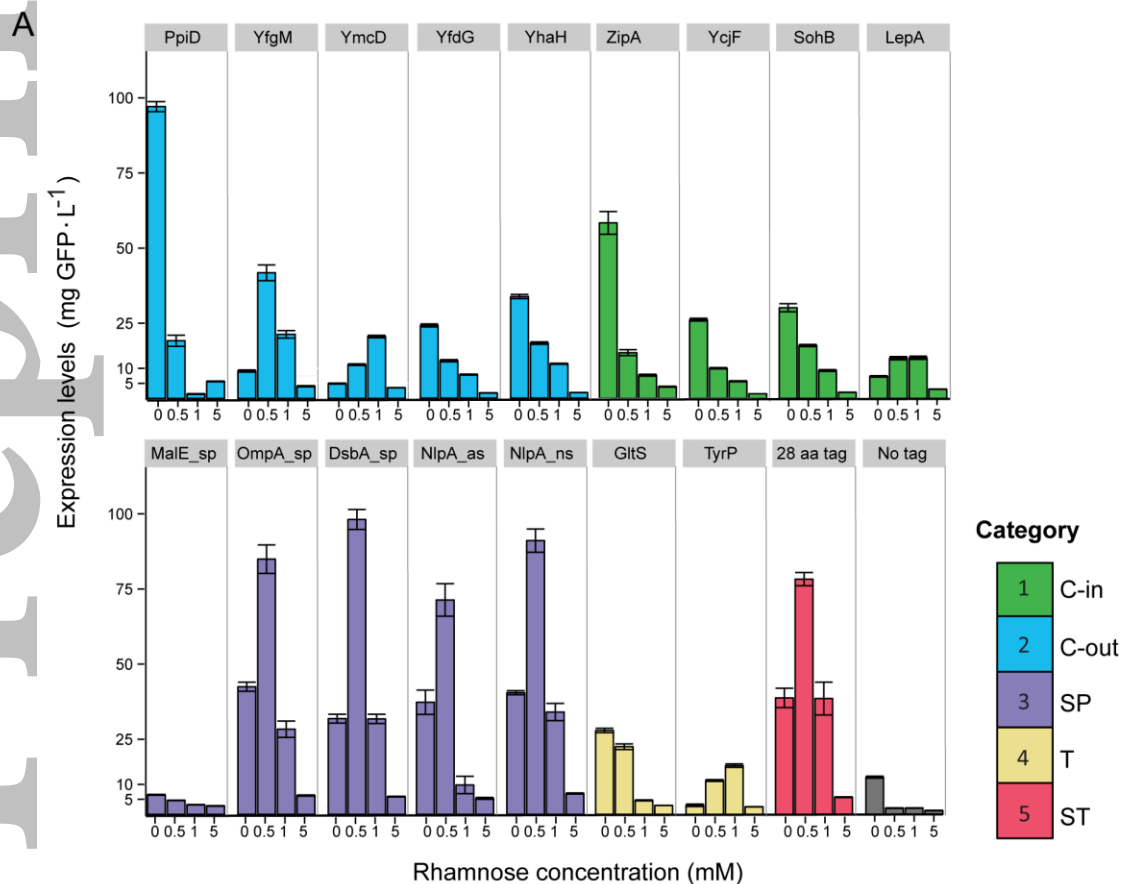
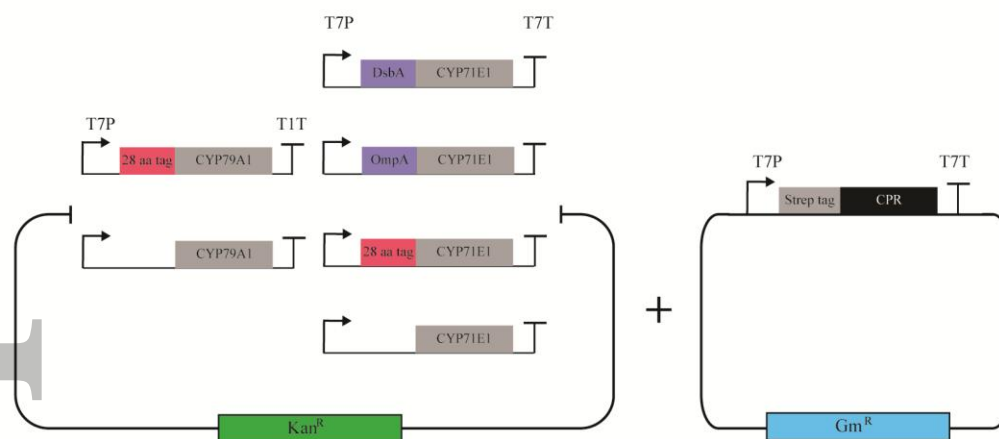
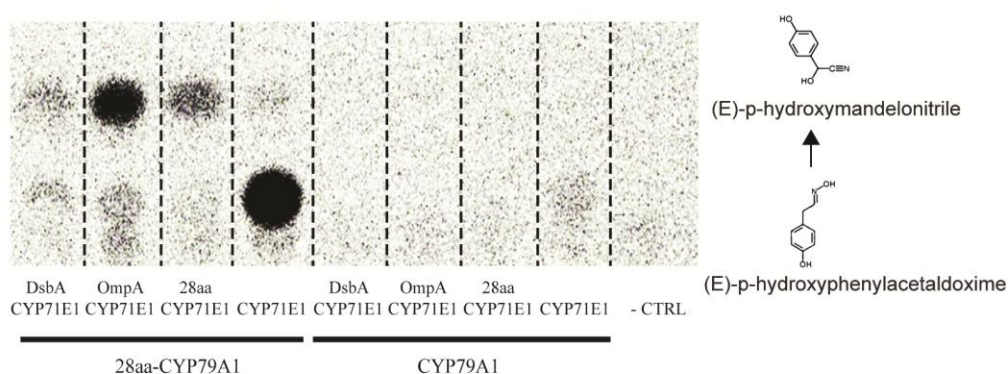


Figure 2

A



B



C

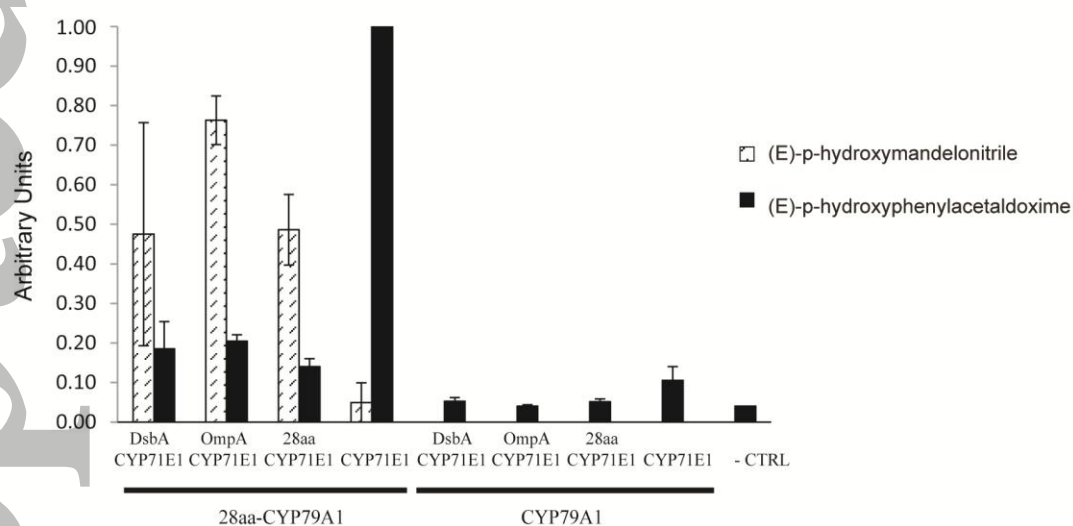


Figure 3

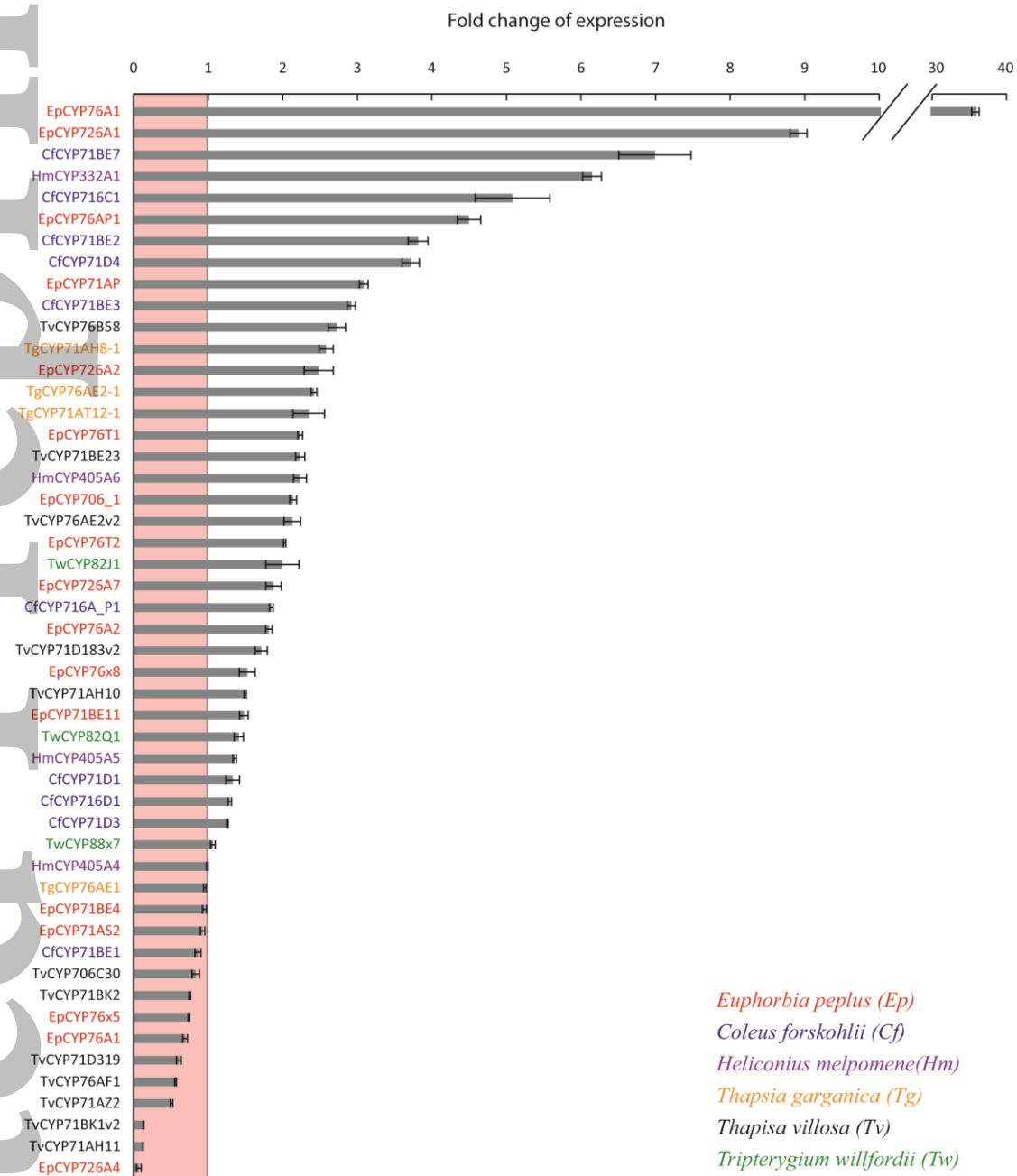


Figure 4